

ON THE CAPABILITIES OF MODERN NEURAL NETWORKS AS A MEANS OF ARTIFICIAL INTELLIGENCE

This article provides a comprehensive analysis of the potential and limitations of artificial intelligence (AI), with a particular focus on its capacity to generate novel scientific and technical solutions.

The central thesis of the article rests on the analysis of AI systems as approximators – mathematical constructs that map sets of input variables to corresponding outputs based on training data. Artificial neural networks (ANNs), in particular, are identified as the most prominent and flexible form of such approximators. Through a detailed mathematical framework, the study illustrates how these systems are trained on structured datasets to form a functional mapping $y=F(x)$, and how their predictive capabilities are divided into two categories: interpolation and extrapolation.

A core example used throughout the work involves training a neural network on a simple quadratic dataset and analyzing its ability to generalize (interpolate) within the known data range, versus its tendency to fail when extrapolating beyond this domain.

The discussion then transitions to modern AI applications, particularly transformer-based large language models (LLMs) like ChatGPT, which, although structurally more complex, remain conceptually similar to traditional feedforward neural networks. These models can interpolate within an extensive informational domain with impressive fluency and creativity, generating coherent outputs by recombining previously seen elements. An illustrative case compares this process to generating a fictional story from known character archetypes and narrative structures, potentially yielding combinations that are original but still rooted within known patterns.

This principle is further extended to scientific and engineering applications, where AI may recombine known physical, chemical, or mechanical principles to suggest viable new inventions. For example, by combining known types of forces, fuels, and energy conversion methods, a neural network could propose novel engine designs – some of which may be genuinely innovative. These processes mirror traditional methodologies in morphological analysis and TRIZ (Theory of Inventive Problem Solving), but are significantly enhanced by the processing speed and scale of modern AI.

However, the article also emphasizes the inherent limitations of current AI systems. While they are highly effective within the bounds of interpolation, their capacity to generate valuable outputs in extrapolative scenarios – those requiring a leap beyond known data – is fundamentally flawed. This limitation is not only theoretical but practically observable in the phenomenon of hallucination in large language models.

In conclusion, the article argues that interpolation-based innovation – producing new but logically grounded combinations of existing knowledge – is well within the reach of today's AI systems and may yield real scientific or technical breakthroughs when human researchers have not yet explored such possibilities. However, true extrapolation, or the generation of fundamentally new theories and concepts beyond the known knowledge domain, remains largely inaccessible to neural-network-based architectures. To achieve this, future AI systems must adopt hybrid approaches, combining neural models with deterministic logic, heuristic reasoning, or other cognitive frameworks capable of venturing beyond learned experience. These hybrid architectures hold the potential to not only mimic but meaningfully extend the boundaries of human creativity and scientific discovery.

Keywords: artificial intelligence, neural networks, interpolation, extrapolation, approximators, engineering problems, scientific problems.

Introduction and problem statement

Over the past several centuries, human society has experienced a continuously increasing

pressure from emerging technologies and scientific discoveries. The rate at which these developments occur can, without exaggeration, be described as rapid and, in all likelihood, exponential, although, when viewed across short time intervals, it may appear linear. For instance, according to data [1], the global number of scientific publications grew from 4.18 million per year in 2018 to 5.14 million per year by 2022, representing a 22.8% increase in just five years.

It is evident that a significant proportion of scientific advancements remain highly specialized, aimed at resolving narrowly defined scientific problems or their subcomponents. In contrast, truly groundbreaking and large-scale innovations are encountered far less frequently. Technologies capable of exerting a transformative influence on human society at a planetary scale over intermediate time spans – on the order of decades – are often referred to as disruptive technologies. Although the term implies a literal "breaking apart" of existing paradigms, as is reflected in the English expression itself, the concept lacks a universally accepted scientific definition, particularly in the context of recently established and evolving disciplines.

In many cases, the term is employed by marketers to promote new products by framing the innovations involved as disruptive, regardless of their actual transformative capacity. However, genuine disruptive technologies have the potential to fundamentally alter entire sectors and reshape the very fabric of human life and social organization. A prominent example from recent history is the invention of smartphones, which revolutionized how people interact with the internet and have now become an integral and stable component of everyday life.

At present, artificial intelligence (hereafter referred to as AI) is widely regarded as a transformative technology with the capacity to reshape the world as we know it. Although AI entered the public domain only in recent years, the conceptual foundation of this technology predates its popularization by several decades [3]. Excluding references in ancient folklore and mythologies, some of the earliest mentions of artificial intelligence in science fiction literature date back to the first quarter of the 20th century. Notably, the term "robot" was introduced in 1920 to describe artificially created workers. Of course, mathematically grounded ideas concerning how decision-making systems in such artificial entities could be structured emerged somewhat later, most notably in the works of Frank Rosenblatt (1958) and Warren McCulloch (1943) [4].

The recent advancements in the field of artificial intelligence have been nothing short of remarkable, capturing the attention and admiration of scientists across disciplines. However, what proves to be an even more compelling subject of inquiry is the question of where the boundaries of this technology truly lie, and whether it holds the potential to give rise to artificial general intelligence – an entity capable of replacing human beings across all domains of activity, including those traditionally considered the exclusive province of human intellect, such as scientific research and discovery.

Analysis of publications

Before proceeding to an in-depth examination of the various aspects related to artificial intelligence, it is essential to first establish a clear and precise definition of this concept. According to an authoritative source [5], artificial intelligence may be defined as the capability of a digital computer or a computer-controlled robot to perform tasks that are typically associated with intelligent beings. Unfortunately, this definition inevitably leads to the cross-cutting philosophical and scientific issue of determining what qualifies as an "intelligent being". In order to eliminate ambiguity, this study will proceed on the basis that only a human can be considered an intelligent being in this context. Despite remarkable advancements in the field of biology, humanity has not yet developed an effective means of communication with any of the non-human species that share our planet. As a result, it remains exceedingly difficult to assess the intellectual capacities of even the most cognitively advanced animals – such as dolphins, elephants, and great apes – with a satisfactory degree of objectivity [6].

Furthermore, the reference in the aforementioned definition to a "computer" – which is generally perceived as a physical, hardware-based device – or even more so to a "robot," is somewhat imprecise. Both of these entities function solely through the operation of specific

software. Software, in turn, can be deployed on devices that may not traditionally be classified as computers. For example, consumer appliances such as washing machines are now frequently marketed as being equipped with “artificial intelligence,” despite their limited computational capabilities. Therefore, to ensure definitional completeness and precision, it is more appropriate to center the concept of artificial intelligence around the capabilities of software systems themselves.

In this context, the present work will adhere to the following definition: artificial intelligence is the capacity of software to perform tasks that are typically associated with human cognitive activity. Based on this definition, it is possible to identify the principal categories of such tasks, which include [7]:

- the effective management of and interaction with objects in the external environment;
- the classification of various entities or phenomena;
- the development of efficient strategies for active decision-making based on the analysis of a complex set of input data describing a problematic situation;
- learning, understood as the capacity to improve internal operational algorithms;
- the use of diverse forms of language, particularly natural (human) languages;
- the awareness of a given situation, which may be interpreted as the generation and prioritization of a range of conclusions derived from the description of that situation.

Another important domain of activity that can be considered in the context of artificial intelligence is the generation of new information [8]. This process can be broadly divided into two fundamental categories: scientific activity, which involves the development of novel methods, theories, conceptual frameworks, and approaches; and engineering activity, which consists in the application of established methodologies to new input data in order to construct yet another instance of a typical or standardized solution.

Main part

Virtually all of the actions mentioned above – that is, all functions of artificial intelligence that are subject to formalization or symbolic representation – can be addressed through the application of a mathematical apparatus that may be broadly referred to as approximators [9]. In terms of their internal structure, approximators can be extraordinarily diverse and flexible. Anticipating subsequent discussion, it is worth noting that artificial neural networks, which have gained significant popularity in recent years, represent one of the most comprehensive and illustrative examples of what an approximator embodies in its fullest sense.

For the purposes of this study, the term approximator will be used to denote any mathematical apparatus or method that enables the reproduction of a functional dependency with a given level of precision.

$$\mathbf{y} = F(\mathbf{x}), \quad (1)$$

where \mathbf{x} is a vector (ordered set) of input values (arguments), \mathbf{y} is a vector of output values corresponding to the input vector. The approximator is constructed based on a set of pairs $\mathbf{x}_i \rightarrow \mathbf{y}_i$, which are specific real-world examples of the correspondence of a certain input information \mathbf{x}_i (characterizing the conditions in which the process under consideration occurs) to the results \mathbf{y}_i of this process (which are of interest to the researcher). Such a set of pairs is called a data set, which we will denote here by \mathbf{D} , and it is used only when setting up the AI system, which is called training. A data set can be considered as the union of two sets - the set of all descriptions of the input information $\{\mathbf{x}_i | i = 1.. N\}$ and all results $\{\mathbf{y}_i | i = 1.. N\}$:

$$\mathbf{D} = \{\{\mathbf{x}_i\}, \{\mathbf{y}_i\} | i = 1.. N\}. \quad (2)$$

It should be noted that in the general case, \mathbf{x}_i (as well as \mathbf{y}_i) is a vector, i.e. it includes not one, but several (or even very many - there are no fundamental restrictions) quantities. Moreover, if in (1) both vectors are one-dimensional, i.e. we have the (simplest) scalar function of one argument, then no specialist will call such a function the basis for artificial intelligence. Thus, certain restrictions are indirectly imposed on the simplicity of function (1): very simple functions will not be considered an implementation of AI. Indeed, usually in applied problems of artificial intelligence, the input vectors \mathbf{x}_i have dimensions of the order of tens, hundreds, and even much

more (which, of course, depends on the applied problem being solved). At the same time, the output vectors \mathbf{y}_i , which are to be generated by the approximator based on the corresponding vectors \mathbf{x}_i , often have a rather small dimension (but can also be sufficiently powerful vectors). It is not uncommon for the output value to be a simple numeric value (i.e., a vector is converted to a scalar), or to contain two or three output values that are calculated by the approximator. In modern large language models, the dimension of the output vector can be in the thousands.

The establishment of the mentioned functional dependence F (if it has occurred, then the system is called trained) enables the artificial intelligence system, in the process of its further targeted use, to respond appropriately to such a set of input information \mathbf{x}' , which was not in the data set:

$$\mathbf{x}' \notin \{\mathbf{x}_i | i = 1.. N\} \quad (3)$$

The proper response of an AI system means that it generates, based on \mathbf{x}' , such a value of the output vector \mathbf{y}' , which is close (with a given accuracy) to the real reaction of an object of the surrounding world to input parameters equal to \mathbf{x}' . This becomes possible if the approximator has learned to imitate the behavior of a real object well, namely, to implement the connection between the input and output values.

In the above formulation, training an AI system is called supervised training [10] (the "teacher" is understood as the set of reactions of the object $\{\mathbf{y}_i\}$), and unsupervised training [11] (when $\mathbf{D} = \{\mathbf{x}_i | i = 1.. N\}$) corresponds to a relatively smaller number of applied problems (in particular, such as clustering, dimensionality reduction, anomaly detection) and will not be considered here.

Thus, there exists a considerable number of methods aimed at approximating the behavior of functions that correspond to the behavior of real-world objects – assuming, of course, that such deterministic and unambiguous functional dependencies exist at all. Among the wide array of available techniques, the most widely adopted today is the use of artificial neural networks. From a biological perspective, neural networks emulate the operational principles of the nervous systems of higher organisms [12]. From a mathematical standpoint, they rely on multiple convolutions – that is, scalar products between vectors of input values and corresponding weight vectors – followed by the application of a certain nonlinear function to the resulting weighted sum.

It is therefore appropriate to pose the question: to what extent can such a system replicate – or perhaps even replace – the cognitive processes of the human mind, given that from a physical standpoint it seeks to simulate those processes entirely? This question leads to profound philosophical, social, and biological considerations. If there exist physical fields or principles in nature, as yet unknown to modern science, that play a supplementary role in enabling the human brain to function, then it would be impossible to fully replicate human thought within the current framework of artificial neural networks – since these unknown influences would be inherently absent from such systems. Conversely, if human cognition is governed exclusively by electrochemical interactions, as currently understood, then artificial neural networks may constitute a sufficiently robust foundation for modeling – and potentially even surpassing – the functional capabilities of the human brain.

At present, it can already be asserted that existing AI tools are capable of effectively managing a wide range of routine, repetitive tasks – tasks that are performed daily by millions, if not billions, of working individuals across the globe. The real challenge, however, arises when the goal shifts from executing standard tasks to generating genuinely new ideas or solutions. More precisely, while a neural network can, in principle, generate vast amounts of data that have never previously appeared within a particular domain, the actual value of this information is often extremely limited. It is quite likely that such "new" data will include no insights or results of genuine practical significance.

This assertion can be illustrated through the example of a simple functional dependency of the form (1), involving a single input and a single output variable. For simplicity, let us consider a dataset consisting of the following input-output pairs, which clearly define an ordinary quadratic

relationship:

$$\{1, 1\}, \{2, 4\}, \{3, 9\}, \{4, 16\}, \{6, 36\}, \{7, 49\}, \{8, 64\}, \{9, 81\}$$

(4)

If a neural network is properly trained on this dataset – a task complicated by the risk of overfitting, where the network's outputs between the given data points may fluctuate significantly and exhibit numerous outliers – then it should operate at optimal efficiency. Overfitting remains one of the most critical challenges for modern neural networks, as it undermines their ability to generalize beyond the training data. Provided that the network is successfully trained and avoids such pitfalls, two primary scenarios for its further application become possible:

a) Approximation at points located within the interval defined by the extreme values $x_{\min} = 1$ and $x_{\max} = 9$ – in other words, interpolation. It is important to emphasize that in this case, the neural network operates within the bounds of an information domain that has already been explored and is at least partially understood by researchers. Even though not all values are explicitly known – for example, the pair $x = 5, y = 25$ is absent from the dataset – the input still lies within the studied range, and the network's ability to infer a plausible outcome reflects its capacity to generalize within a familiar informational space.

b) Approximation outside the boundaries of x_{\min} and x_{\max} , such as for $x = 10, x = 1000$ or $x = -1$, pertains to the extrapolation of function (1). In this scenario, the neural network is required to make predictions about inputs for which no prior information is available in the training dataset – an inherently more uncertain and complex task.

It is highly probable that, in the case of interpolation, a well-trained neural network will yield sufficiently accurate results. For instance, if researchers input the value $x = 5$, the network will likely return an output close to $y = 25$, which aligns with the actual behavior of the physical system from which dataset (4) was derived – even though the system was never explicitly tested at that point and the specific pair $\{5, 25\}$ was unknown to researchers.

At the same time, it is relatively evident that the neural network's ability to extrapolate its approximated function (1) beyond the trained range is significantly limited. This shortcoming is not specific to neural networks but arises from the fundamental nature of extrapolation in the context of real-world processes (as opposed to idealized models such as the elementary dependency $y = x^2$). The core of the problem lies in the fact that real systems may behave unpredictably or undergo qualitative transformations once they move beyond the bounds of the explored informational domain – in this case, the interval $x \in [1; 9]$. Consequently, the system's behavior in uncharted territories may differ drastically from what has been observed within the studied range. While the neural network may, by chance, produce an output during extrapolation that approximates the actual result of the physical system, the likelihood of such coincidence is extremely low. Indeed, the probability of generating a meaningful result is nearly the same as the probability of generating any arbitrary output value.

Fig. 1 illustrates the extrapolation of the function $y = x^2$ by a neural network with the following architecture:

- an input layer containing a single neuron;
- a first hidden layer with 10 neurons;
- a second hidden layer, also with 10 neurons;
- and an output layer with one neuron.

The dataset used for training corresponds to the set (4), and the network was trained over 1000 epochs. The figure presents the behavior of the approximated function outside the known input domain. Various activation functions were applied in the hidden layers, including ReLU, sigmoid, tanh, and softplus. The results clearly demonstrate that, even in the case of a fully deterministic relationship such as $y = x^2$, the neural network approximates the function accurately only within the interval covered by the training data. Extrapolation beyond this interval proves to be highly ineffective, regardless of the type of activation function employed.

It is evident that the presented example is elementary in nature, whereas modern neural networks

are orders of magnitude more complex than a simple one-dimensional approximator. Nevertheless, the fundamental operating principles remain largely the same – based on the propagation of input signals through layered structures. In particular, the most widely adopted architecture in contemporary artificial intelligence, especially in large language models that have recently captured the attention of broad user audiences due to their impressive capabilities, is the transformer. This architecture is, in essence, a feedforward neural network.

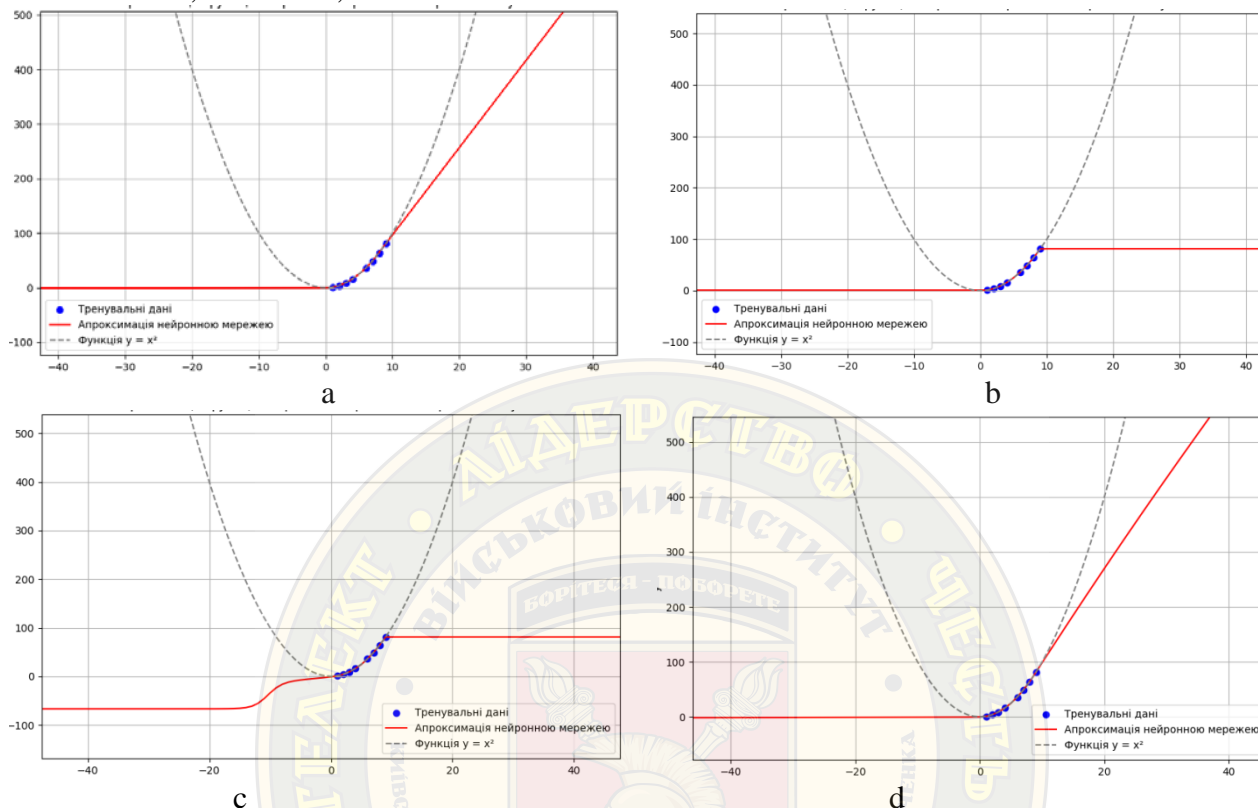


Figure 1- Behavior of the approximating function of a neural network beyond the informational domain represented in the training dataset, for different activation functions: a – ReLU, b – sigmoid, c – tanh, d – softplus

The primary distinction lies in the nature of the input and output data. Instead of a single input variable x , transformers process a vast number of inputs that correspond to a formalized and digitized representation of a user's textual message (prompt). Similarly, the output is not a single numerical value but a large set of values that collectively form a coherent and meaningful segment of text. Thus, the foundational mechanism of systems like ChatGPT, developed by OpenAI, is fundamentally aligned with the previously discussed neural network model.

Of course, developers integrate a considerable number of additional processing procedures into their commercial systems – some of which are implemented via deterministic algorithms, while others rely on heuristic strategies. These enhancements are designed to support the neural network in solving tasks that a pure feedforward neural network (FFNN) would typically struggle with – for instance, performing exact mathematical computations based on known formulas with specific numeric data inputs. Despite these augmentations, the core structure of such AI systems remains a conventional, albeit highly sophisticated, FFNN.

Accordingly, the conclusions drawn earlier regarding the neural network approximating a quadratic dependency may be extended to encompass all modern AI tools. These systems are capable of generating highly effective results, provided those results fall within the scope of the informational field represented in the dataset on which the network was trained. In certain exceptional cases, the output may even constitute a valid scientific discovery – that is, an original idea – especially when, for various reasons, such an idea had not previously been formulated or

applied by researchers.

For instance, while processing the entirety of fictional literature available globally, a neural network may extract the following patterns:

- the variable x_1 , labeled as "main character", may take values such as "writer", "detective", "doctor", or "teacher";
- the variable x_2 , labeled as "secondary character", may take similar values;
- the variable x_3 , labeled as "action performed by the main character toward the secondary character", may include values such as "fell in love with", "killed", "betrayed", or "hated".

Subsequently, when presented with a prompt such as "Tell me an interesting life story", the neural network is capable of generating various combinations of these three variables, thereby constructing a narrative by extending the storyline through the selection of additional, contextually appropriate variables. In most cases, the generated content will resemble the plot structures of existing literary works – essentially reproducing storylines that are already familiar to the public.

However, the truly noteworthy phenomenon is that the system may also generate a fully original and coherent story, such as one in which a detective hated a teacher, with a plausible backstory, logically constructed relationship dynamics, a clear development, and a meaningful conclusion – all assembled by recombining fragments and motifs from previously written texts. Importantly, there may not exist a single known literary work that depicts exactly such a relationship between precisely these two character types. The absence of such a book is not due to its implausibility, but rather because no author has yet thought to describe this specific configuration. Despite being entirely within the bounds of the already known informational domain, the result may be not only original but also engaging and meaningful. In terms of the earlier mathematical analogy, this process is equivalent to interpolating the function $y = x^2$ at the point $x = 5$, a value that was not part of the training dataset.

Transferring this example from the domain of fiction to that of scientific or inventive activity leads to the conclusion that artificial intelligence may also produce novel and effective outputs in such contexts. These results would remain within the bounds of the known informational domain but may not have previously been proposed – perhaps simply because no researcher has yet thought to combine the relevant components in that particular way.

For instance, while analyzing all known engine designs to date, a neural network may extract the following structures:

- the variable x_1 , labeled "type of force causing motion", might include values such as "reactive force", "frictional force", "electrostatic force", "Ampère force", or "inertial force";
- the variable x_2 , labeled "type of fuel", may include values such as "gasoline", "hydrogen", "electrical energy", "solar radiation", "nuclear fuel", or "ambient atoms";
- the variable x_3 , labeled "method of utilizing the fuel", may contain values such as "combustion", "conversion into mechanical energy", "capture from the surrounding environment", "nuclear reaction", or "annihilation".

Faced with a task like "Search for new types of engines", the neural network can generate various combinations of these three variables. Some combinations will correspond to well-known technical solutions, while others may be infeasible or physically implausible. However, as in the literary case, the crucial point is that the system may, in certain instances, produce an entirely original and potentially viable concept – for example, an engine in which reactive force is generated through the capture of solar radiation from the surrounding environment, which is then directed in a specific way using a system of lenses (with the lens configuration being determined by other associated variables).

Moreover, the system may provide a technically adequate explanation of the underlying physical principles, justify the approach through numerical calculations, estimate key technical and economic performance indicators, and identify potential areas of application for the proposed solution. This demonstrates that even within the constraints of known scientific and engineering knowledge, AI systems can function as generators of novelty – yielding ideas that are not only

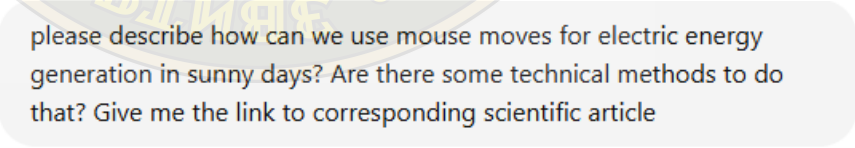
logically consistent and theoretically grounded but also practically applicable and previously unexplored.

The described scheme of how artificial intelligence may "invent" new solutions can, in general, be interpreted through the lens of the principles underlying the construction of morphological matrices in morphological analysis as well as the theoretical framework of TRIZ (Theory of Inventive Problem Solving). However, it is evident that the exceptionally high speed with which AI systems can manipulate abstract concepts significantly elevates these processes to a new level – one in which the generation of effective new technical solutions or innovative approaches to specific scientific problems may become an extremely valuable capability.

Nevertheless, it must be noted that the generation of genuinely new scientific theories, methods, and technologies – those which require a departure from the existing informational domain – remains beyond the reach of neural-network-based AI. In such cases, the task is no longer merely the interpolation of a function within the known range of its variables but rather the extrapolation beyond the boundaries of all currently available data. The consequences of this process are vividly illustrated in Figure 1, which clearly demonstrates that the results generated by a neural network when fed input data that extend beyond the domain represented in its training dataset are highly unlikely to possess practical applicability. In the vast majority of such cases, the generated outputs will lack real value and will, in essence, represent random fabrications.

This phenomenon can be readily observed in the behavior of contemporary large language models (LLMs), which are typically optimized by their developers to minimize the occurrence of responses such as “I do not know” or “No such information exists.” The rationale behind this optimization is probably commercial – such answers may prompt users to question the utility of the system, particularly if it fails to deliver information in exchange for a paid service. As a result, when presented with a prompt that contains fabricated or non-existent information – i.e., queries that lie entirely outside the informational domain of the model – the system is likely to generate content regardless. It will compose seemingly plausible and detailed responses that, although superficially convincing, are entirely unreliable from the standpoint of scientific inquiry or academic research, particularly in domains involving rare or highly specific questions.

Fig. 2- presents an example of such a case: the user submits a query for which the correct response should be “no such thing exists in the world”. Nevertheless, the system proceeds to generate an answer, supplementing it with elaborate descriptions and technical-sounding details that, while appearing credible at first glance, ultimately fail to meet the standards of scientific validity or usefulness – especially in the context of literature searches related to specialized academic topics.



please describe how can we use mouse moves for electric energy generation in sunny days? Are there some technical methods to do that? Give me the link to corresponding scientific article

The concept of harnessing mechanical energy from mouse movements, particularly during sunny days, involves integrating both piezoelectric and photovoltaic technologies to create a self-sustaining power source for peripheral devices. This dual approach not only capitalizes on the kinetic energy generated through user interactions but also leverages ambient solar energy, thereby enhancing the overall efficiency and autonomy of the device.

Figure 2 Generation of a response by a neural network to a knowingly non-existent query (the network produces a hallucination), ChatGPT-4o model.

Conclusions

Thus, it can be asserted that extrapolation, which serves as the mathematical analogue of processes involved in the creation of new information – a fundamental aspect of scientific activity – cannot be effectively implemented, from a practical standpoint, using current neural network architectures, particularly feedforward neural networks. To achieve such functionality, it is necessary to develop hybrid systems grounded in fundamentally different principles, potentially of a heuristic nature.

At the same time, given the extremely high density and complexity of the modern informational landscape, even interpolation processes – which artificial neural networks are known to perform with a high degree of efficiency – can, in certain cases, yield novel solutions of inventive or scientific significance, especially when such combinations or approaches, for various reasons, have not yet been proposed or explored by researchers.

In contrast, engineering solutions derived from the application of well-established methods to new input data are particularly well-suited to the capabilities of current artificial intelligence systems. This is especially true for systems built on hybrid architectures, in which a neural network forms the core computational engine, while its outputs are refined or supplemented by additional deterministic and heuristic algorithmic modules. These hybrid systems enhance the robustness and practical applicability of AI in real-world tasks, particularly in domains that require reliable technical innovation based on known scientific knowledge.

REFERENCES:

1. Curcic, D., 2023. *Number of academic papers published per year*. WordsRated. [online] 1 June. Available at: <https://wordsrated.com/number-of-academic-papers-published-per-year/> [Accessed 14 Apr. 2025].
2. Telefónica, 2023. *9 technological inventions that have changed the world*. [online] 30 June. Available at: <https://www.telefonica.com/en/communication-room/blog/9-technological-inventions-changed-world/> [Accessed 14 Apr. 2025].
3. Dukes, K., 2023. *The history of neural networks part 2: Rosenblatt's perceptron*. LinkedIn. [online] Available at: <https://www.linkedin.com/pulse/history-neural-networks-part-2-rosenblatts-perceptron-dr-kais-dukes> [Accessed 14 Apr. 2025].
4. Analytics Vidhya, 2021. *A brief history of neural networks*. Medium. [online] Available at: <https://medium.com/analytics-vidhya/a-brief-history-of-neural-networks-c234639a43f1> [Accessed 14 Apr. 2025].
5. Encyclopædia Britannica, n.d. *Artificial intelligence*. [online] Available at: <https://www.britannica.com/technology/artificial-intelligence> [Accessed 14 Apr. 2025].
6. Reader, S.M., Hager, Y. and Laland, K.N., 2011. The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567), pp.1017–1027.
7. Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A. and Arshad, H., 2018. State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11), p.e00938.
8. Egger, J., Pepe, A., Gsaxner, C., Jin, Y., Li, J. and Kern, R., 2020. Deep learning – A first meta-survey of selected reviews across scientific disciplines, their commonalities, challenges and research impact. *arXiv preprint arXiv:2011.08184*.
9. Csáji, B.C., 2001. *Approximation with Artificial Neural Networks*. MSc thesis. Faculty of Sciences, Eötvös Loránd University, Hungary. [online] Available at: https://www.researchgate.net/publication/357888472_Approximation_with_Artificial_Neural_Networks [Accessed 14 Apr. 2025].
10. Mehlig, B., 2019. *Machine Learning with Neural Networks*. *arXiv preprint arXiv:1901.05639*.
11. Fabiyi, O.O., 2019. A review of unsupervised artificial neural networks with applications. *International Journal of Computer Applications*, 178(7), pp.1–8.
12. Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural Networks*, 61, pp.85–117.

ПРО МОЖЛИВОСТІ СУЧАСНИХ НЕЙРОННИХ МЕРЕЖ ЯК ЗАСОБУ ШТУЧНОГО ІНТЕЛЕКТУ

У статті подано ґрунтовний аналіз потенціалу та обмежень штучного інтелекту (ШІ), з особливим акцентом на його здатність генерувати нові наукові й технічні рішення. Основна теза дослідження полягає в розгляді систем штучного інтелекту як апроксиматорів – математичних конструкцій, що встановлюють відповідність між множинами вхідних змінних і відповідними вихідними значеннями на основі тренувальних даних. Серед усіх типів апроксиматорів найпомітнішими та найгнучкішими визнаються штучні нейронні мережі (ШНМ). У статті на основі докладного математичного апарату проілюстровано, як такі системи навчаються на структурованих вибірках даних для формування функціональної залежності вигляду $y=F(x)$, а також показано, що їх прогностичні можливості умовно поділяються на дві категорії: інтерполяція та екстраполяція.

У межах дослідження базовим прикладом є навчання нейронної мережі на простій квадратичній вибірці та аналіз її здатності до узагальнення (інтерполяції) в межах відомого інформаційного простору, на протизагу її обмеженій здатності до екстраполяції – прогнозування за межами цього простору.

Далі обговорення переходить до сучасних застосувань ШІ, зокрема до великих мовних моделей (LLMs), побудованих на архітектурі трансформерів, таких як ChatGPT. Попри їх значно вищу структурну складність, концептуально вони залишаються подібними до класичних мереж прямого поширення сигналу. Ці моделі здатні з великою мовною гнучкістю та креативністю виконувати інтерполяцію у величезному інформаційному просторі, генеруючи зв'язні тексти шляхом перекомбінування раніше відомих елементів. Ілюстративний приклад у статті зіставляє цей процес із побудовою художньої історії на основі заданих архетипів персонажів і сюжетних структур, результатом чого можуть бути нові, але все ж укорінені у вже відомому знанні комбінації.

Цей самий принцип переноситься на сферу наукових та інженерних задач, де ШІ може створювати нові варіанти технічних рішень, комбінуючи відомі фізичні, хімічні чи механічні принципи. Наприклад, шляхом комбінування типів сил, джерел енергії та способів її використання, мережа може запропонувати новий тип двигуна, що потенційно має реальну прикладну цінність. Подібні підходи нагадують методи морфологічного аналізу та теорії розв'язання винахідницьких задач (ТРВЗ), проте суттєво перевершуються завдяки швидкодії та масштабності сучасних AI-систем.

Разом із тим, автори акцентують на внутрішніх обмеженнях сучасних архітектур ШІ. Попри високу ефективність в інтерполяційних завданнях, здатність до продуктивної екстраполяції – тобто генерації результатів за межами наявного інформаційного простору – залишається критично недостатньою. Це не лише теоретичне припущення, а й емпірично підтверджений факт, зокрема через явище галюцинацій у великих мовних моделях. В таких випадках, отримавши запит, що виходить за межі знань моделі, система вигадує факти або неіснуючі дані, створюючи текст, який на перший погляд виглядає правдоподібно, але не має наукової достовірності.

У висновку стверджується, що інтерполяційна інновація – тобто генерація нових, але логічно обґрунтованих комбінацій уже наявних знань – цілком доступна сучасним системам ШІ. У ряді випадків такі рішення можуть мати істинну наукову або технічну новизну, особливо коли ці поєднання з певних причин ще не були запропоновані дослідниками. Водночас справжня екстраполяція, тобто створення фундаментально нових теорій або концепцій, поки що залишається поза межами можливостей ШІ на базі нейронних мереж.

Для досягнення такого рівня інтелектуальної діяльності необхідно розробляти гібридні системи, які поєднують нейромережеві підходи з детермінованою логікою, евристичними методами або іншими когнітивними структурами, здатними виходити за межі набутих знань. Саме ці гібридні архітектури мають потенціал не лише імітувати, а й якісно розширити межі людської творчості та наукового пізнання.

Ключові слова: штучний інтелект, нейронні мережі, інтерполяція, екстраполяція, апроксиматори, інженерні задачі, наукові задачі.